

An extended formalism for preferential attachment in heterogeneous complex networks

A. Santiago* and R. M. Benito†

Grupo de Sistemas Complejos, Departamento de Física y Mecánica,
Escuela Técnica Superior de Ingenieros Agrónomos,
Universidad Politécnica de Madrid, 28040 Madrid, Spain.

(Dated: Received February 2, 2008)

In this paper we present a framework for the extension of the preferential attachment (PA) model to heterogeneous complex networks. We define a class of heterogeneous PA models, where node properties are described by fixed states in an arbitrary metric space, and introduce an affinity function that biases the attachment probabilities of links. We perform an analytical study of the stationary degree distributions in heterogeneous PA networks. We show that their degree densities exhibit a richer scaling behavior than their homogeneous counterparts, and that the power law scaling in the degree distribution is robust in presence of heterogeneity.

PACS numbers: 89.75.Fb, 89.75.Hc, 89.75.-k

A complex network is a set of nodes and links with a non-trivial topology [1]. In the current effort to achieve a single coherent framework for complex systems, network theory has focused on the underlying principles that govern their topology [2, 3]. Dynamical network models [4] are stochastic discrete-time dynamical systems that evolve networks by the iterated addition/subtraction of nodes/links. These models regard network topology as an emergent property of the network evolution, focusing on the mechanisms that take place on such process. Among these mechanisms, preferential attachment (PA) [5] enjoys a foremost position in network literature.

The PA model by Barabási and Albert [6] has provided a minimal account of mechanisms for the emergence of *scale-free* networks [7, 8]. Such networks are characterized by a degree distribution according to a power law, $P(k) = k^{-\gamma}$, which leads to a non-negligible presence of *hubs* (highly connected nodes). The PA model assumes two mechanisms: growth and preferential attachment. The process starts with a seed and a new node is added to the network at each step. Each new node has a number m of links attached, which are connected to the existing nodes following the so-called *attachment rule*: the linking probability of a network node v_i is proportional to its degree k_i , $\Pi(v_i) = k_i / \sum_j k_j$. This step is iterated until a number N of nodes have been added.

The PA model is strictly topological as the node degrees $\{k_i\}$ are the only metrics that drive network evolution. Nevertheless, the assumptions implicit in the PA model are not valid for a wide class of complex systems. Often, interactions between individual elements are mediated by their intrinsic properties. Although network theory has led to a significant improvement in our understanding of complex systems, it has been argued that its framework should be augmented in order to improve our

modelling of complexity [1, 9]. We will refer to *heterogeneous networks* as networks where node intrinsic properties induce affinities in their interactions. We consider such networks as a logical first step in addressing the complication introduced by the influence of individual elements on the network structure.

In recent years several dynamical models have incorporated the influence of element properties. These include weighted networks [10, 11, 12], social models [13, 14], competition models [15, 16], a local knowledge model [17], a metric model [18], a gas-like model [19] and network automata [20]. In this letter we propose a general class of heterogeneous PA models where a function measures the *affinity* between node intrinsic properties (*states*) and biases the attachment probabilities of links in a generalized PA rule. The introduction of affinities aims to provide a more realistic modelling of the structure of numerous real systems that exhibit biased interactions. The proposed class contains previous models as particular cases, and provides a framework for the systematic analysis of the influence of heterogeneity in PA networks.

The PA model can be easily generalized to heterogeneous networks by imposing a metric structure on the node states, while preserving the basic mechanisms of growth and preferential attachment of the original Barabási-Albert model. Next we formally define a general class of heterogeneous PA models:

Definition 1 A *heterogeneous PA model with global affinity* M_1 is a 3-tuple (R, ρ, σ) , where: (1) R is an arbitrary metric space. The elements $x \in R$ are the node **states**. (2) $\rho : R \mapsto \mathbb{R}^+ \cup \{0\}$ is a nonnegative function with unit measure over R referred as **node state distribution**. (3) $\sigma : R^2 \mapsto \mathbb{R}^+ \cup \{0\}$ is a nonnegative function referred as **affinity** of the interactions.

The class of heterogeneous models with global affinity (HPA_g) is the set of all 3-tuples that satisfy the conditions in Definition 1. This formalism defines the evolution of a network according to the following rules:

*antonio.santiago@upm.es

†rosamaria.benito@upm.es

- (i) The nodes v_i are characterized by their state $x_i \in R$. The node states describe intrinsic properties deemed constant in the timescale of evolution of the network.
- (ii) The growth process starts with a seed composed by N_0 nodes (with arbitrary states $x_i \in R$) and L_0 links.
- (iii) A new node v_a (with m links attached) is added to the network at each iteration. The number m is common for all the added nodes and remains constant during the evolution of the network. The newly added node is randomly assigned a state x_a following the distribution $\rho(x)$.
- (iv) The m links attached to v_a are randomly connected to the network nodes following a distribution $\{\Pi(v_i)\}$ given by an extended *attachment rule*,

$$\Pi(v_i) = \frac{\pi(v_i)}{\sum_j \pi(v_j)}, \quad \pi(v_i) = k_i \cdot \sigma(x_i, x_a). \quad (1)$$

The visibility π of a node v_i in the attachment rule is given by the product of its degree k_i and its affinity σ with the newly added node v_a , which is itself a function of the states x_i and x_a . It thus can be seen that for each interaction σ biases the degree k_i of the candidate node. Steps (iii) and (iv) are iterated until a desired number of nodes has been added to the network. To sum up, the choice of the tuple (R, ρ, σ) determines the form of heterogeneity in the attachment mechanism.

Next we derive an analytical solution for the stationary degree distribution $P(k)$ of the class HPA_g of heterogeneous models. The solution is obtained by *rate equations* which establish a balance in the flows of degree densities over a partition of a growing network. Let us first define a sequence of functions $\{f(k, x, N)\}_{N>0}$ which measure the probability density of a randomly chosen node having degree $k \in \mathbb{N}$ and state $x \in R$ in a network at the iteration $t = N$. The degree densities are local metrics, thus they uniformly converge when $N \rightarrow \infty$ to a stationary density function $f(k, x)$. Finally, the stationary degree distribution $P(k)$ measures the probability of a randomly chosen node having degree k in the thermodynamic limit.

The rationale for the choice of f in modelling the evolution of the degrees is simple: in a heterogeneous model, nodes with equal degree may contribute differently to changes in $P(k, N)$ according to their state, despite being homogeneously accounted by P . Thus, for each (k, N) we need to know the influence of x on the degree distribution. We will denote by $V(x) = \{v_i, x_i = x\}$ the subset of nodes in the network with state x . Assuming that the assignation of states x_i is uncorrelated with the topology of the growing network, and that there are no linking events between existing nodes, the sequence $\{f\}$ can be modelled on each $V(x)$. For each x , the form of the equation will be $L_1 - L_2 = R_1 - R_2$, where:

L_1 = density of nodes with degree k at $t = N + 1$;

L_2 = density of nodes with degree k at $t = N$;

R_1 = increase in density due to nodes with degree $k - 1$ that have gained a link at $t = N$;

R_2 = decrease in density due to nodes with degree k that have gained a link at $t = N$.

The total density of nodes with state x and degree k at $t = N + 1$ is $(N + 1)f(k, x, N + 1)$. Here we assume that N is large and no incoming node is rejected due to lack of affinity, therefore we approximate the network size by t . Likewise, the total density of nodes with state x and degree k at $t = N$ is $Nf(k, x, N)$. In order to estimate the changes in f due to the acquisition of new links, we introduce a mean-field *fitness* factor $w(x)$ that measures the average affinity of a network node with state x towards the incoming nodes,

$$w(x) \equiv \int_R \sigma(x, y) \rho(y) dy. \quad (2)$$

The probability of a node being chosen for attachment is proportional to $k w(x)$, and the decrease in the density of nodes with state x and degree k due to their attachment of a new link is proportional to $k f(k, x, N) w(x)$. This term must be normalized by the increase in all the densities (irrespective of their degree k and state x) in the network at $t = N$, measured by a partition factor

$$\psi(N) \equiv \sum_k k \int_R f(k, x, N) w(x) dx. \quad (3)$$

As in each iteration there are m links attached to a newly added node, the decrease in the total density of nodes with state x and degree k due to nodes with degree k gaining a link and being promoted to degree $k + 1$ is $m k f(k, x, N) w(x) / \psi(N)$ at each step. Here we assume that for large N the probability that a single node receives more than one link attached to the same new node becomes negligible. Likewise, the increase in the total density of nodes with state x and degree k due to nodes with state x and degree $k - 1$ gaining a link and being promoted to degree k is $m(k - 1) f(k - 1, x, N) w(x) / \psi(N)$. The resulting density rate equation for $k > m$ is:

$$(N + 1)f(k, x, N + 1) - Nf(k, x, N) = \frac{m w(x)}{\psi(N)} [(k - 1)f(k - 1, x, N) - k f(k, x, N)]. \quad (4)$$

When $k = m$, there is no increase in $f(m, x, N)$ due to the promotion of nodes with $k = m - 1$. However, each newly added node has degree m and this contribution to $f(m, x, N)$ is distributed according to $\rho(x)$, thus:

$$(N + 1)f(m, x, N + 1) - Nf(m, x, N) = \rho(x) - \frac{m w(x)}{\psi(N)} m f(m, x, N). \quad (5)$$

There are no nodes with degree $k < m$, since when $N \rightarrow \infty$ all the links attached to newly added nodes find receptive nodes in the network, therefore the previous equations define all the possible cases in each iteration.

In the thermodynamic limit $N \rightarrow \infty$, $f(k, x, N + 1) = f(k, x, N) = f(k, x)$ and the rate equations become:

$$f(k, x) = \begin{cases} \frac{m w(x)}{\psi} [(k - 1)f(k - 1, x) - k f(k, x)] & \text{for } k > m, \\ \rho(x) - \frac{m w(x)}{\psi} m f(k, x) & \text{for } k = m, \end{cases} \quad (6)$$

These equations are coupled by the stationary partition factor $\psi \equiv \sum_k k \int f(k, x) w(x) dx$. Notice that in the homogeneous PA model, $w(x) = 1 \forall x \in R$ and thus $\psi = \sum_k k P(k)$, which is equal to the average network degree $z = (2Nm + 2L_0)/(N + N_0)$. In the stationary limit $N \gg N_0$ and $Nm \gg L_0$, thus $\psi \simeq 2m$.

To the contrary, in the heterogeneous models $w(x)$ continuously weighs the density $f(k, x)$ in the integral associated to each degree k in the stationary partition factor ψ . In order to decouple Eq. 6, let us assume that $w(x)$ doesn't change too much on R so that we may approximate $w(x)$ by its mean $\bar{w} \equiv \int_R w(x) \rho(x) dx$, then

$$\psi \simeq \sum_k k \int f(k, x) \bar{w} dx = \sum_k k P(k) \bar{w} \simeq 2m \bar{w}. \quad (7)$$

The resulting decoupled system becomes:

$$\begin{aligned} f(k, x) = & \quad (8) \\ \left\{ \begin{array}{ll} \frac{1}{2} \left(\frac{w(x)}{\bar{w}} \right) [(k-1)f(k-1, x) - kf(k, x)] & \text{for } k > m, \\ \rho(x) - \frac{1}{2} \left(\frac{w(x)}{\bar{w}} \right) m f(k, x) & \text{for } k = m, \end{array} \right. \end{aligned}$$

We introduce a *normalized fitness* $\hat{w} = w(x)/\bar{w}$, which measures the fitness of a state relative to the average fitness of the network nodes. Solving Eq. 8 we obtain

$$\begin{aligned} f(m, x) &= \frac{2\rho}{\hat{w}m + 2}, \\ f(k, x) &= \frac{\hat{w}(k-1)}{\hat{w}k + 2} f(k-1, x) \text{ for } k > m. \end{aligned} \quad (9)$$

Thus the solution for the stationary density for $k > m$ is

$$f(k, x) = \left(\prod_{j=m+1}^k \frac{\hat{w}(j-1)}{\hat{w}j + 2} \right) \frac{2\rho}{\hat{w}m + 2} \quad (10)$$

and integrating the density in Eq. 10 over the state space R the stationary degree distribution is

$$P(k) = \int_R \left(\prod_{j=m+1}^k \frac{\hat{w}(j-1)}{\hat{w}j + 2} \right) \frac{2\rho}{\hat{w}m + 2} dx. \quad (11)$$

It is important to remark that the solutions in Eqs. 10 and 11 are valid for any model in our class, since they do not make any assumption regarding the geometry of space R , the form of affinity σ or the distribution ρ .

Furthermore we may assume that $w(x) > 0 \forall x \in R$ without loss of generality, otherwise $\sigma(x_0, y) = 0 \forall y \in R$ since $\sigma \geq 0$, which would mean that x_0 is "inert" and may be excluded from R . Then $\hat{w}(x) > 0 \forall x \in R$ and we may write the density in Eq. 10 for $k > m$ as

$$f(k, x) = \frac{2\rho/\hat{w}}{m + 2/\hat{w}} \frac{B(k, 1 + 2/\hat{w})}{B(m, 1 + 2/\hat{w})}. \quad (12)$$

where Legendre's Beta function $B(y, z) = \int_0^1 t^{y-1} (1-t)^{z-1} dt$ for $y, z > 0$ satisfies the functional relation

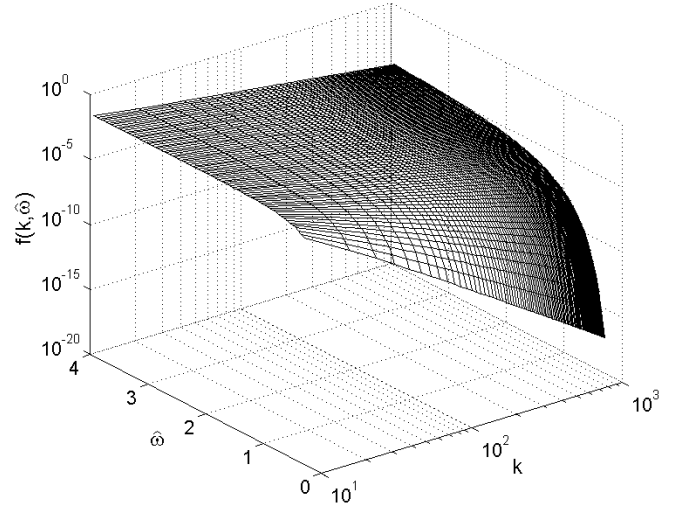


FIG. 1: Solution for the stationary density $f(k, \hat{w})$ in heterogeneous PA models (Eq. 10) with $\rho(x) = 1$ and $m = 3$.

$\Gamma(a)/\Gamma(a+b) = B(a, b)/\Gamma(b)$ for Euler's Gamma Γ function. Likewise, integrating the density in Eq. 12 we obtain for $k > m$ the stationary degree distribution

$$P(k) = \int_R \frac{2\rho/\hat{w}}{m + 2/\hat{w}} \frac{B(k, 1 + 2/\hat{w})}{B(m, 1 + 2/\hat{w})} dx. \quad (13)$$

Notice that Eq. 12 reduces to $f(m, x)$ in Eq. 9 when $k = m$, thus Eqs. 12 and 13 are valid for $k \geq m$. Also, notice that the stationary densities $f(k, x)$ in the heterogeneous formalism, as given by Eq. 12, follow the form of a Beta function with arguments $(k, 1 + 2/\hat{w})$.

In particular cases it is possible to obtain simpler expressions for $f(k, x)$. First, in the homogeneous case, $w(x) \equiv 1$, thus $\hat{w} = w = 1$ and Eq. 10 yields

$$f(k, x) = \frac{2\rho m(m+1)}{k(k+1)(k+2)}, \quad (14)$$

and the density functions $f(k, x)$ asymptotically follow power laws $f(k, x) \sim k^{-\gamma}$ with exponent $\gamma = 3$ irrespective of the node states. This is reasonable, as the dynamics of the homogeneous PA network is not affected by them, thus the asymptotic behavior of all density components is the same. If furthermore we assume that $\rho(x)$ is uniform over R , then any subset $V(x)$ of nodes will exhibit the same distribution as the whole network. Likewise, the degree distribution $P(k)$ asymptotically follows a power law with exponent $\gamma = 3$ like the Barabási-Albert model [8], irrespective of the distribution ρ .

In presence of heterogeneity the fitness $w(x)$ is generally not constant over R . It is still possible to obtain simple forms of $f(k, x)$ from Eq. 10 for discrete values of \hat{w} that render the denominator a product of integers. For instance, let x be such that $\hat{w} = 2 > 1$, then

$$f(k, x) = \frac{\rho m}{k(k+1)}, \quad (15)$$

which asymptotically behaves as a power law with exponent $\gamma = 2$. Conversely, letting x be such that $\hat{w} = 2/3 < 1$, then Eq. 10 yields

$$f(k, x) = \frac{3\rho m(m+1)(m+2)}{k(k+1)(k+2)(k+3)}, \quad (16)$$

which asymptotically behaves as a power law with exponent $\gamma = 4$. By extension, normalized fitness values $\hat{w} = 2/n$ with integer n yield density functions $f(k, x)$ with an asymptotic power law behavior with $\gamma = 1 + 2/\hat{w}$.

More generally, it can be proved that the Beta function behaves as $B(y, z) \sim y^{-z}$ when $y \rightarrow \infty$. Therefore, Eq. 12 implies that $f(k, x)$ behaves asymptotically as $B(k, 1 + 2/\hat{w}) \sim k^{-(1+2/\hat{w})}$. The density components $f(k, x)$ in heterogeneous PA networks exhibit a *multiscaling* according to power laws $k^{-\gamma(\hat{w})}$ along the continuous spectrum of normalized fitness \hat{w} . The scaling exponents are inversely related to \hat{w} according to $\gamma(\hat{w}) = 1 + 2/\hat{w}$ so that they span themselves a continuum.

Thus, as \hat{w} increases (resp. decreases), the exponent γ of f decreases (resp. increases). Nodes with states more fit than the average ($\hat{w} > 1$) adopt densities with $\gamma < 3$, which exhibit a slower asymptotical decay, and tend to produce more hubs. Nodes with states less fit than the average ($\hat{w} < 1$) adopt densities with $\gamma > 3$, which exhibit a faster asymptotical decay. These results are verified in Fig. 1, which shows the plot of the density f . Although the shape of f depends on the particular heterogeneity of the network, its asymptotic behavior can be characterized in a general way by plotting the dependence of f on $\hat{w}(x)$ instead of x .

The behavior described points out to a *signature* of heterogeneity in the topology of PA networks: the fitness $w(x)$ will tend to fluctuate on R and the density components $f(k, x)$ will exhibit different scaling exponents. This prompts us to suggest a procedure for empirically detecting such signature in real networks. In order to check the role of a certain degree of freedom in the network evolution, one could select different states $\{x\}$ varying only in such dimension, select subsets of nodes $V_\delta(x) = \{v_i, d(x^{(i)} - x) < \delta\}$ with states in a small neighborhood of the states x . If the partial distributions $\{P(k|x)\}$ exhibit power laws with different exponents, it would be feasible to assume this dimension (or another one correlated with it) is biasing the attachment mechanism. To check the existence of heterogeneity it would be enough to find one such subset of nodes whose partial distribution $\{P(k|x)\}$ exhibits a power-law with an exponent different to the one in $P(k)$.

As for the degree distribution $P(k)$ of the heterogeneous models, Eq. 11, it is obtained by the integration of the density components which form a set of power-laws with varying exponents. The level of relative variability present in the fitness $w(x)$ over the state space R will

determine the ultimate composition of $P(k)$. Choices of (R, ρ, σ) that yield little fluctuations in $w(x)$ on R will translate into density functions $f(k, x)$ deviating slightly from the homogeneous power law, and degree distributions similar to the Barabási-Albert case. On the other hand, choices of (R, ρ, σ) that yield larger fluctuations in $w(x)$ will translate into a wider spectrum of exponents γ in the density components, and thus a larger deviation of the degree distribution from the homogeneous case.

While we need to know the tuple (R, ρ, σ) to obtain the distribution $P(k)$, it is still possible to make rough predictions about its behavior in heterogeneous PA networks. We may regard the degree distribution in such networks as a superposition of power laws whose exponents are distributed around the homogeneous $\gamma = 3$. For low degrees it can be expected that averaging low and high γ components will tend to produce a behavior of $P(k)$ similar to the one observed in homogeneous networks. When $k \rightarrow \infty$, the tail of $P(k)$ will be dominated by the slowest decaying components, associated to states with highest w . Furthermore, given that \hat{w} will be distributed by definition around 1, the $P(k)$ of all the models in the class will exhibit scaling exponents satisfying $1 < \gamma \leq 3$. This is remarkable since the empirical observations in real scale-free networks show that most exponents fit within this interval.

In summary, we have proposed a generalization of the PA model to heterogeneous networks. We have defined a general class of heterogeneous PA models where node properties are described by fixed states in an arbitrary space, and where an affinity function biases the attachment probabilities of links. We have obtained analytical solutions for the degree densities and degree distribution of heterogeneous networks in the stationary limit. We have shown that these networks exhibit a richer scaling behavior than their homogeneous counterparts, a multiscaling of their degree densities according to power laws with exponents spanning a continuum. Such phenomenon leaves a signature of heterogeneity in the topology of PA networks that can be empirically checked in real networks. We have also shown the relationship between the variability in the exponents and the fluctuations in the fitness over the state space, and the consequences of this on the asymptotic behavior of the degree distribution. Finally we have shown that the degree distributions of all heterogeneous models in our class exhibit power laws with exponents within the limits empirically observed in real networks. These results suggest that the ubiquity of real scale-free networks with otherwise different structural properties may be partially explained by PA mechanisms with varying levels of heterogeneity.

This work has been supported by the Spanish MEC under Project 'i-MATH' No. CSD2006-00032 and Project No. MTM2006-15533, and GESAN, S.A.

-
- [1] S. H. Strogatz, *Nature* **410** 268 (March 2001).
 - [2] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74** 47 (2002).
 - [3] M. E. J. Newman, *SIAM Review* **45** 167 (2003).
 - [4] S. N. Dorogovtsev and J. F. F. Mendes, *Adv. Phys.* **51** 1079 (2002).
 - [5] D. J. S. Price, *Science* **149** 510 (1965).
 - [6] A.-L. Barabási and R. Albert, *Science* **286** 509 (1999).
 - [7] A.-L. Barabási, R. Albert and H. Jeong, *Phys. A* **272** 173 (1999).
 - [8] S. N. Dorogovtsev, J. F. F. Mendes and A. N. Samukhin, *Phys. Rev. Lett.* **85** 4633 (2000).
 - [9] L. A. N. Amaral and J. M. Ottino, *Eur. Phys. J. B* **38** 147 (2004).
 - [10] S.-H. Yook, H. Jeong, A.-L. Barabási and Y. Tu, *Phys. Rev. Lett.* **86** 5835 (2001).
 - [11] A. Barrat, M. Barthelemy and A. Vespignani, *Phys. Rev. Lett.* **92** 228701 (2004).
 - [12] M. Barthelemy, A. Barrat, R. Pastor-Satorras and A. Vespignani, *Phys. A* **346** 34 (2005).
 - [13] M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera and A. Arenas, *Phys. Rev. E* **70** 056122 (2004).
 - [14] A. Grabowski and R. A. Kosinski, *Phys. Rev. E* **73** 016135 (2006).
 - [15] G. Bianconi and A.-L. Barabási, *Europhys. Lett.* **54** 436 (2001).
 - [16] G. Ergün and G. Rodgers, *Phys. A* **303** 261 (2002).
 - [17] J. Gómez-Cardenes and Y. Moreno, *Phys. Rev. E* **69** 037103 (2004).
 - [18] D. J. B. Soares, C. Tsallis, A. M. Mariz and L. R. da Silva, *Europhys. Lett.* **70** 70 (2005).
 - [19] S. Thurner and C. Tsallis, *Europhys. Lett.* **72** 197 (2005).
 - [20] R. Alonso-Sanz, *Chaos* **32** 1285 (2006).